3. THE TRITUS SYSTEM

Submitting/Adding/Giving natural language questions (e.g. , â€œ How do I tie shoelaces ? â€) to search engines in their original form often does not work very well . Search engines typically retrieve documents similar to the original queries . Unfortunately , the documents with the best answers may contain only one or two terms from the original queries . Such useful documents may then be ranked low by the search engine , and will never be examined/viewed by typical users who do not look beyond the first page of results . To answer a natural language question , a promising approach is to automatically reformulate the question into a query that con- tains terms and phrases that are expected to appear in documents containing answers to the original question .

3.1 Problem Statement

We focus on the first step of the question answering process : retrieving a set of documents likely to contain/provide an answer to a given question . These documents are then returned as the output of the system . The returned documents can be examined by a human user directly , or passed on to sophisticated answer extrac- tion modules of a question answering system (e.g. , Abney et al . [2000] , Mann [2002] , Prager et al . [2002] , and Radev et al . [2002]) . Thus , it is crucial that the answer to a question of interest be present in this set of initially retrieved docu- ments . At the same time , the set of retrieved documents can not be so large that it overwhelms the user or the subsequent (typically computationally expensive) answer extraction components . Therefore , our goal is to return a reasonable- sized set of documents that , at the same time , must contain an answer to the question . We now formally state/acknowledge/define the problem that we are addressing .

3.2 Learning to Transform Questions into Effective Queries

We attempt/sought/seek to find transformations from natural language questions into ef- fective queries that contain terms or phrases expected to appear in documents that contain answers to the question . Our learning process is shown in Figure 2 .

3.2.1 Selecting Question Phrases.

In the first stage of the learning process (Step (1) in Figure 2) , we generate a set of phrases that identify different cat- egories of questions where the questions in each category have a similar goal . For example , the question â€œ What is a hard disk ? â€ implies that the user is look- ing for definitions or descriptions of a hard disk . The goal of the question can be inferred from the question phrase â€œ what is a â€ .